DOCUMENT RESUME

ED 234 747 IR 010 813

AUTHOR Kemerer, Richard W.; Schmid, Richard F.
TITLE Evaluating a Complex Program: Where to Start and How to Finish.
PUB DATE [83]
NOTE 22p.; Paper presented at the annual Conference of the National Society for Performance and Instruction (Detroit, MI, April 10-15, 1983).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; *Data Analysis; *Distance Education; Evaluation Methods; *Formative Evaluation; Higher Education; Instructional Materials; *Material Development; *Program Evaluation
IDENTIFIERS *Adjusted Agreement Index; Canada

ABSTRACT

Data on the intrinsic characteristics of an educational program are essential in pilot testing new programs in order to determine how successfully design components have achieved their intended purpose. At the outset of an evaluation of such data, it is necessary to define the evaluator's role and determine what intrinsic information will be required, when it will be needed, who can provide it, and how it will be collected. The Adjusted Agreement Index (AAI) was used in a case study to summarize data from a complex program developed by the Institute of Canadian Bankers, in which eight university-level distance education courses were simultaneously pilot-tested for 26 weeks with over 1,000 Canadian students. The AAI score depicts relative agreement by respondents and is calculated by subtracting the percentage of respondents who disagree with a statement from the percentage who agree. Results indicate that the AAI is a helpful tool in summarizing results clearly so that decision-makers will know which intrinsic characteristics require immediate attention for revision purposes. The AAI is easy to calculate and use with subjective response questionnaires. Two references are listed, and four figures illustrate evaluation design, AAI calculation, score distribution for various AAI results, and student perceptions of assignments by course. (LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
[x] This document has been reproduced as received from the person or organization originating it.
[ ] Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED234747

EVALUATING A COMPLEX PROGRAM: WHERE TO START AND HOW TO FINISH

Lengthy questionnaires are frequently administered when assessing instructional materials. However, evaluators typically find themselves overwhelmed with data, and use only key questions in an "eyeball" fashion. This presentation offers a case study of training evaluation, and methods utilized to overcome data interpretation problems. A unique "Adjusted Agreement Index" is discussed as an effective means of evaluating subjective responses with large samples. These techniques are useful for clear presentations to superiors, and content revision using questionnaires.

Richard W. Kemerer, Ph.D.
Manager, Program Development
Bank of Montreal
26 Wellington Street East
Toronto, Ontario
M5E 1S2

Work: (416) 867-8315
Home: (416) 762-0565

Richard F. Schmid, Ph.D.
Associate Professor
Concordia University - Montreal
1455 Ouest, Boul. de Maisonneuve
Montréal, Quebec
H3G 1M8

Work: (514) 879-8108
Home: (514) 489-7322

IR010813

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Richard W. Kemerer
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EVALUATING A COMPLEX PROGRAM: WHERE TO START AND HOW TO FINISH

Dr. Richard Kemerer is Manager, Program Development, Bank of Montreal. Previously, Richard was Assistant Director, Institute of Canadian Bankers, where he directed research and development projects. Richard's interests are in developing training programs which are compatible with principles of adult education.

Dr. Richard Schmid is Associate Professor, Educational Technology, Concordia University. Richard is presently conducting research in areas of algorithms in instructional design and cognitive processes.

EVALUATING A COMPLEX PROGRAM: WHERE TO START AND HOW TO FINISH

Often evaluators are called upon to provide information on the intrinsic characteristics of an education program. These non-performance data are essential in pilot testing new programs in order to determine how successfully design components have achieved their intended purpose (Scriven, 1967). The purpose of this paper is to explore some of the key questions to be answered in evaluating intrinsic data and to illustrate how the Adjusted Agreement Index (AAI), which was developed by these authors, was used to summarize data in an easily presentable fashion.

This study focuses on a complex program developed by the Institute of Canadian Bankers. Eight university level courses were pilot tested simultaneously: Accounting, Business Administration, Business Finance, Business Strategy, Communications, Economics, Marketing and Organizational Behaviour. Each course was designed by a different professor. Four Canadian universities were represented. All courses were designed for distance education and all materials were administered directly from the Institute's head office in Montreal through nine Canadian Banks. Over 1,000 students throughout Canada were involved in the 26-week pilot test.

WHERE TO START AN EVALUATION

1. What is the evaluator's role? A key decision at a study's outset is to identify the role of the evaluator. By definition, evaluation must reach a decision. What part will the evaluator play in making that decision? In our case, it was

decided that the evaluator's role was to provide information rather than make decisions based on that information. This arrangement thus follows Stufflebeam's Context, Input, Process, Product approach to evaluation. (Stufflebeam, 1968)

2. What type of decision is needed? A second consideration concerns the type of evaluation decisions which are to be made. As our example was a pilot-test of a new program, decisions would be required on how to revise the program based upon how well the design led to its intended outcome. Thus, the focus was on identifying design flaws and delivery problems, constituting formative, rather than summative evaluation.

3. What intrinsic information is required? A third consideration concerns information. In our situation, the study focused on finding out who the students were, how well they linked up with the content, and how well the content itself was structured for effective instruction. Therefore, three information types were identified: personal data, delivery system, and teaching material.

4. When will the information be required? A fourth consideration is timing. In our case, the greatest initial concern was on how well students linked up with content. Thus, information on the success of the delivery system was required almost as soon as the program was initiated. More specifically, were students receiving material on time? What condition were materials in upon arrival? Information on teaching materials was required at interim points throughout

the 26-week term before various decisions were to be made regarding assignments, revisions to lesson plans, or in interpretation of student difficulty with textbooks.

5. Who can provide information? A fifth consideration concerns sources of information. In our case, four sources were identified: assignment correctors, program administrators, dropouts, and students sitting final exams. Assignment correctors were professors from the various universities contracted by the Institute to grade assignments. Program administrators assisted in the delivery of materials through various banks in Canada. They were located in regional offices of Canadian banks, at the Institute's regional offices and at their head office in Montreal. Dropouts were defined as students who did not complete the first two assignments. Students sitting final exams were those who actually made it through the 26 weeks and had registered for the final exam.

6. How will the information be collected? A sixth consideration focuses on data collection. Two different types of information gathering instruments were identified: questionnaires, and small group interviews (focus groups). Questionnaires were felt to be appropriate in obtaining information from assignment correctors, dropouts, and students sitting the final exams. With reference to dropouts, the questionnaire could be distributed directly by assignment correctors because they had the most up-to-date class lists and information on which students had not turned in their

assignments. The other two groups could be contacted directly by the Institute. Focus group interviews were felt to be appropriate for gathering information from program administrators because of their small numbers, their geographical diversity and the need for immediate information for decision-making.

The overall evaluation design is represented in Figure 1. This particular design was agreed upon before the first student enrolled in the program because it was felt that if all contributors to the program assisted in designing the evaluation study, they would have a vested interest in acting on evaluation results. Therefore, program administrators, course designers, evaluators, and instructors all met in a "strategy session" to map out information collection methods, time frames and information sources.

INSERT FIGURE 1 HERE

HOW TO FINISH

Evaluations frequently produce at least some of the above types of data, but the results are seldomly used properly. Even more often, results are employed in an ad hoc fashion to support decisions made while waiting for the data. The index described in this section is an attempt at making the collected data "worth waiting for", i.e., interpretable.

In this case study, information from the assignment collectors and program administrations was transparent and largely idiosyncratic. Dropouts provided useful systemic data, but is not dealt with here as it does not address the formative question of instructional effectiveness. The following comments are thus restricted to questionnaire results from students sitting final exams. The intent is to introduce the Adjusted Agreement Index (AAI) which was developed to summarize data in an easily presentable fashion.

In our example, 311 final exam questionnaires were completed out of 529, for a 59% return rate. The questionnaires centred on intrinsic data-exploring the effectiveness of the overall course, textbooks, lesson guides, and assignments. The case does not distinguish between those who successfully completed their exams and those who did not. Given these data, the evaluator's challenge was to summarize clearly in order to facilitate administrators' understanding.

Fifty-five statements were followed by 7-point Likert-type scales. Positions on the scale ranged from Strongly Agree, Neither Agree nor Disagree (neutral or no opinion), to Strongly Disagree.

The AAI depicts relative agreement by respondents and is calculated by subtracting the percentage of respondents who disagree with a statement (checked 6 or 7 on the scale) from the percentage who agree (checked 1 or 2). This difference is the AAI score. In Figure 2, 30% strongly agreed with the statement, 20% strongly disagreed, and 50% neither agreed nor disagreed. The AAI is calculated to be 10.

INSERT FIGURE 2 HERE

As illustrated in Figure 3, the AAI is tied to the distribution of responses. In the first example, the distribution is skewed towards the agreement side of the scale. The result is a high positive AAI, indicating greater overall agreement with the statement. In the second example, the distribution

is skewed toward the right, yielding a high negative AAI. This translates as greater overall disagreement with the statement. Examples 3 and 4 indicate normal and bi-modal distributions, both of which yield a low positive or negative AAI. This lower AAI score represents lesser overall agreement or disagreement with the statement and in essence says that most respondents are neutral or are mixed in their opinions.

INSERT FIGURE 3 HERE

A simple decision making rule was established for purposes of identifying components requiring the earliest attention: priority decisions are indicated by relatively high deviations from desirable AAI scores.

Figure 4 illustrates how the decision rule was applied to course assignments. The statement which said that there were an adequate number of assignments yielded a considerable variety of AAI scores. Business Finance, with a score of 70 indicates that a high percentage of students agreed with the statement (100 would be perfect, though highly unlikely). On the other hand, both Accounting and Business Administration yielded AAI scores of 18, indicating considerably less agreement over the number of assignments. The priority decision was to determine whether Accounting and Business Administration required additional assignments.

INSERT FIGURE 4 HERE

As another example, statements which dealt specifically with overall structure were pooled. Great care must be taken to combine only statements addressing the exact same issue. All the courses revealed relatively high AAI scores, varying from 43 to 61. This could be interpreted to mean that the overall structure of assignments was considered to be satisfactory as none deviated significantly from the desired level, and no immediate action was required.

Five statements on the questionnaire dealt with ease in completing assignments. Students in all eight courses indicated their disagreement as reflected in the negative AAI scores. Economics, with an AAI of -42, is the one course where students had the greatest difficulty completing assignments. Accounting, with an AAI of -1, suggests the least difficulty in completing assignments. In this instance, the priority decision was to determine why Economics assignments were perceived as being too hard. Other courses would be addressed in order of the severity of the problem, the number of students effected, the difficulty of revising, and so on.

In our final example, statements concerning the receipt of good feedback yielded both positive and negative AAI scores. Students in Business Finance and Communications indicated through their AAI scores that they received relatively good feedback. Students in Accounting and Business Strategy, on the other hand, indicated they were less happy with feedback. The priority decision was to examine feedback quality in Accounting and Business Strategy.

When AAI scores within a course consistently deviate from an acceptable level, such as in Accounting or Business Administration, that course should be examined more thoroughly. The level of detail provided by the evaluation questionnaire and performance data will further guide revision. The AII is extremely effective at alerting the designer of problems, but must always be considered with other data when available. For example, if the Economics assignments were very difficult, but produced excellent performance results and generally satisfied students (the "pain" is necessary for learning), then revision is less critical, or perhaps even undesirable.

The AAI is thus a relative score based on distribution. It allows direct and easy comparison of each aspect measured. The comparison can be made both with the ideal AAI level, and with other courses/programs where available. One weakness of the AAI is that it is not an absolute score, (a low score on one aspect may be a high score on another aspect), thus potentially confusing to non-users. However, direct comparisons should dissipate such problems. The AAI is also not very useful if the quantity of data is small. However, for a complex program such as the example described above, the AAI successfully distinguished areas requiring immediate attention. Finally, the use of the AAI requires the evaluator to conscientiously attend the to response distribution. A bimodal distribution may produce a nondescript AAI, but suggest important differences within the target population, such as prior knowledge or language. The AAI fails to independently alert the user to this type of problem, but so do all other statistics.

A final word is necessary regarding standard descriptive statistics. The mean, median and mode all reflect central tendencies, or averages, whereas an evaluator is often interested in extremes, or deviations from the norm. Percentages, when left untreated, are even more relativistic and open to interpretation than the AAI. The AAI, in fact, forces the user to recognize both the value of deviation scores, and the limitations inherent in their non-absolute level of interpretation. In other words, they are easy to interpret only when the criterion for establishing a meaning is supplied. This guards against both confusion and misinterpretation.

In summary, intrinsic characteristics of an educational program affect how well the program achieves its intended purpose. After defining the evaluator's role, it is necessary at a study's outset to identify what information will be required, when it will be needed, who can provide the information, and how the information will be collected. The Adjusted Agreement Index is a helpful tool in summarizing results clearly so that decision-makers will know which intrinsic characteristics require immediate attention for revision purposes. This case study involving eight courses designed for distance education suggests that the AAI is easy to calculate and use with subjective response questionnaires.

References

Scriven, M. The methodology of evaluation. In R.E. Stake (Ed.), Curriculum Evaluation. American Educational Research Association monograph series on evaluation, no.1, Chicago: Rand McNally, 1967.

Stufflebeam, D.L. Evaluation as enlightenment for decision-making. Columbus, Ohio: Evaluation Center, Ohio State University, 1968.

FIGURE 1

EVALUATION DESIGN

SOURCE	TIME FRAME	COLLECTION METHOD	INFORMATION		
			PERSONAL DATA	DELIVERY SYSTEM	TEACHING MATERIAL
ASSIGNMENT CORRECTORS	- BY 2ND ASSIGNMENT	QUESTIONNAIRE			X
	- AFTER FINAL EXAM				X
PROGRAM ADMINISTRATORS	- TWO MONTHS INTO TERM	FOCUS GROUP INTERVIEWS		X	X
DROPOUTS	- BY 2ND ASSIGNMENT	QUESTIONNAIRE (DISTRIBUTED BY CORRECTORS)	X	X	X
STUDENTS SITTING FINAL EXAMS	- BY FINAL EXAM	QUESTIONNAIRE	X	X	X
	- AFTER FINAL EXAM	FOCUS GROUP INTERVIEWS			X

FIGURE 2

THE ADJUSTED AGREEMENT INDEX (AAI)

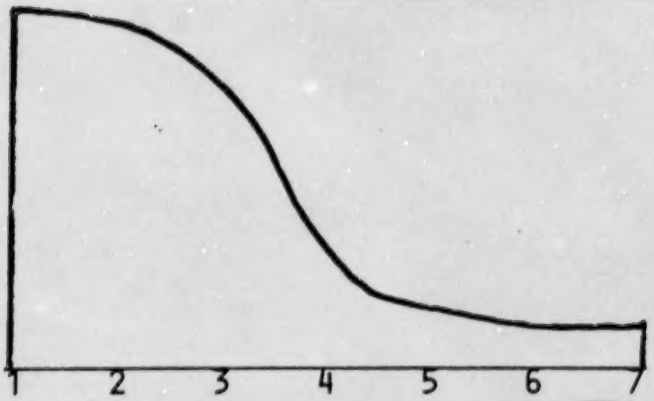
WHERE DO THE NUMBERS COME FROM?

	STRONGLY AGREE		NEITHER AGREE NOR DISAGREE (NEUTRAL OR NO OPINION)		STRONGLY DISAGREE
SCALE:	1 2		3 4 5		6 7
COMPUTATION:	PERCENTAGE WHO AGREE, I.E. CHECK 1 OR 2 ON SCALE	—	PERCENTAGE WHO DISAGREE, I.E. CHECK 6 OR 7 ON SCALE	=	ADJUSTED AGREEMENT INDEX
EXAMPLE:	30% (WHO AGREE)	—	20% (WHO DISAGREE)	=	10 (ADJUSTED AGREEMENT INDEX)*

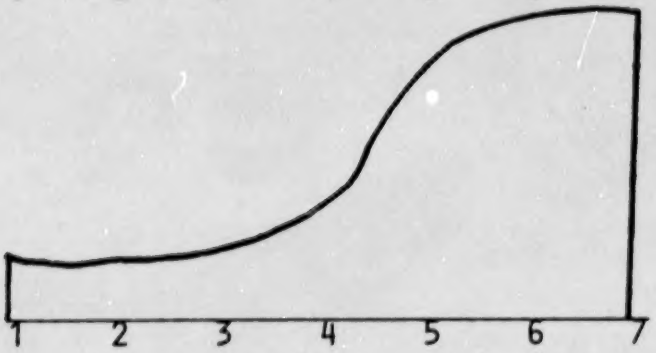
* THE HIGHER THE POSITIVE OR NEGATIVE ADJUSTED AGREEMENT INDEX, THE GREATER THE OVERALL AGREEMENT.

FIGURE 3

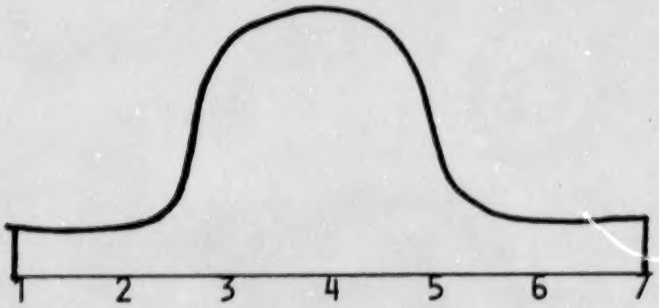
DISTRIBUTION OF SCORES

1. 

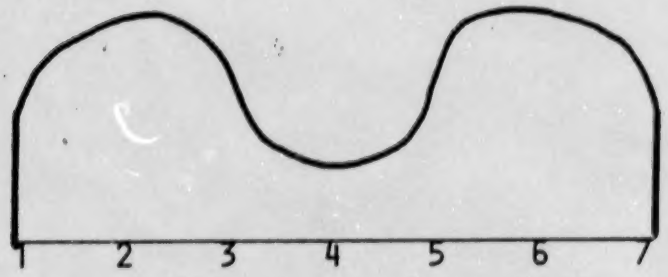
= HIGH POSITIVE AAI

2. 

= HIGH NEGATIVE AAI

3. 

= LOW + OR - AAI

4. 

= LOW + OR - AAI

FIGURE 4

STUDENT PERCEPTIONS OF ASSIGNMENTS BY COURSE

	ACCT.	B.ADMIN.	B.FIN.	B.STRAT.	COMM.	ECON.	MARK.	O.B.
ADEQUATE NUMBER (1 STATEMENT)	18	18	70	62	59	66	46	44
OVERALL STRUCTURE (5 STATEMENTS)	50	50	53	43	49	61	52	56
EASE IN COMPLETING (5 STATEMENTS)	-1	-17	-23	-11	-5	-42	-22	-25
RECEIVED GOOD FEEDBACK (3 STATEMENTS)	-10	4	33	-10	25	-6	-3	1

* THE HIGHER THE NUMBER, THE GREATER THE AGREEMENT